Postgenomics Diagnostics: Metabolomics Approaches to Human Blood Profiling

Oxana Trifonova, Petr Lokhov, and Alexander Archakov

Abstract

We live in exciting times with the prospects of postgenomics diagnostics. Metabolomics is a novel “omics” data-intensive science that is accelerating the development of postgenomics diagnostics, particularly with use of accessible peripheral tissue compartments. Metabolomics involves the study of a comprehensive set of low molecular weight substances (metabolites) present in biological systems. The metabolite profiles represent the molecular phenotype of biological systems and reflect the information encoded at the genomic level and implemented at the transcriptomic and proteomic levels. Analysis of the human blood metabolite profile is a universal and highly promising tool for clinical postgenomics applications because it reflects both the endogenous and exogenous (environmental) factors influencing an individual organism. This article presents a critical synthesis and original analysis of both the technical implementation of metabolic profiling of blood and statistical analysis of metabolite profiles for effective disease diagnostics and risk assessment in the present postgenomics era.

Introduction

BIOLICAL SYSTEMS FUNCTION through complex interactions of four main components: the genome, transcriptome, proteome, and metabolome (Fig. 1). The metabolome, defined as a comprehensive set of low molecular weight compounds (metabolites) of a biological system, is a unique component as the metabolites represent the building blocks for all other biochemical structures, including proteins (amino acids), genes, and transcripts (nucleotides), which form the other “-omes” of biological systems. Metabolites are synthesized within the biological system (endogenous) or can be imported from outside (exogenous). They serve as a source for energy generation, construction, and production of the molecules.

The term “metabolomics” is generally defined as the comprehensive study of all metabolites in the biological system. The study of the metabolome allows us to understand the biological function and phenotype, and it can be applied either alone or in combination with other postgenomics investigations (Atherton et al., 2009; Nicholson and Lindon, 2008; Schnackenberg, 2007). By analyzing the metabolites involved in pathophysiological processes, the information gained from metabolomics can be applied in clinics for developing effective diagnostics and risk assessments of various diseases (Barrett, 2012; Dunn et al., 2011; Schnackenberg and Beger, 2006; Schnackenberg, 2007).

Analytical Platforms for Metabolomics Analysis

In general, metabolomics studies can be divided into two types, depending on experimental workflow and required biological information—a targeted analysis and an untargeted analysis (or metabolic profiling). The first group of studies is usually performed to realize a well-defined hypothesis and to investigate certain targets—small numbers of metabolites that are related to specific biochemical processes or classes. Therefore analytical methods for targeted analysis include extensive separation of analytes and sample matrix and include the construction of calibration curves and quantification of metabolites. However, for many metabolomics studies, only a general hypothesis can be developed. For example, it is known that there are metabolic differences between healthy individuals and those diagnosed with particular diseases, but identification of the metabolites specifically related to pathophysiological changes has not been clear to date. In such cases, it is important to obtain reliable data on numerous metabolites present in multiple classes or metabolic pathways (i.e., to obtain a metabolite profile). Subsequent data analysis can provide novel insights into metabolome changes related to the pathophysiological processes and help to generate a hypothesis on the investigated process. Thus, metabolic profiling is an initial step for subsequent targeted studies (Kelland and Oliver, 2004; Schnackenberg and Beger, 2006).

FIG. 1. A biological system represented as a complex interaction of the genome, transcriptome, proteome, and metabolome.

Metabolic profiling provides information regarding up to thousands of metabolites in a given sample due to the application of high-throughput “global” methods (so-called profiling methods). Today, Nuclear Magnetic Resonance (NMR) Spectroscopy and Mass Spectrometry (MS) are the most frequently used methods for metabolic profiling of biofluids (Dunn et al., 2011; Psychogios et al., 2011).

NMR spectroscopy

NMR spectroscopy is an invaluable tool for chemists and structural biologists, and has been used extensively in metabolic profiling research for more than 20 years. NMR spectroscopy has the benefit of being quantitative, highly reproducible, and, unlike other profiling platforms, nonselective. This means that a relatively large number of different metabolites can be detected simultaneously. The sample preparation for NMR-based metabolomics experiments is simple and nondestructive, and allows conducting several analyses on the same sample. It can even be used to analyze intact plasma/serum samples. The large molecular weight molecules contained in plasma/serum such as phospholipids, triglycerides, and lipoproteins give rise to broad signals in the resulted NMR spectra that may obscure the narrow resonances of small molecular weight metabolites which are often of greater biological interest. To facilitate the observation of narrower resonances, appropriate sequences can be used. For example, the most commonly used method for suppressing the broad signals from large molecules is the multiple pulse spin-echo experiment Carr–Purcell–Mieboom–Gill (CPMG) (Rooney et al., 2003; Smith et al., 2007; Zhang et al., 2010). The biggest disadvantage of NMR spectroscopy compared to other analytical platforms is its low sensitivity. Therefore, NMR spectroscopy can reliably detect and quantify only metabolites present in relatively high concentrations.

Using simple one-dimensional NMR, only several tens of metabolites can usually be detected in blood plasma or serum (Brindle et al., 2002; Kirschenlohr et al., 2006; Stringer et al., 2011) (Fig. 2).

Mass spectrometry

Mass spectrometry has advanced the frontier of metabolomics due to its ability to analyze up to thousands of metabolites in a given sample (Fig. 3). Mass spectrometry can be applied with or without preliminary chromatographic or electrophoretic separation. Direct Infusion (or injection) Mass Spectrometry (DIMS) is used with Electrospray Ionization (ESI)–mass spectrometers when the sample is directly introduced into the mass spectrometer (Vaidyanathan et al., 2002). Since biological samples are highly complex, an instrument with high mass resolution and accuracy is required to ensure fit-for-purpose mass separation of the majority of detected metabolites.

DIMS provides a high-throughput system capable of analyzing up to 60 samples per hour, though with a reduced capacity of definitive metabolite identification and an increased level of ion suppression. The ion suppression observed in ESI—when multiple compounds in complex metabolomics samples are competing for the available charge—influences the quantitative analysis of mass spectra. Therefore, DIMS is rarely used for this purpose (Dunn et al., 2011; Ellis et al., 2007).

To eliminate the disadvantages of DIMS that are typical when analyzing complex samples, a chromatographic or capillary electrophoretic separation is applied. This approach reduces the ion suppression considerably and removes the background noise, thereby enhancing the dynamic range of metabolite detection. In general, two types of chromatographic separation are applied with mass spectrometry: Gas Chromatography (GC) and Liquid Chromatography (LC).

FIG. 2. A NMR spectrum of human blood plasma metabolites. Adapted from Stringer et al., 2011.

Gas chromatography coupled with mass spectrometry (GC-MS) is one of the most often used techniques in metabolomics. GC-MS is commonly used in the Electron Ionization (EI) mode, which is one of the oldest and best established methods for ionizing materials. The largest collection of reliable reference spectra currently available for the identification of chemicals including metabolites is provided by the Electron Ionization (EI) Library, which is part of the NIST/EPA/NIH Mass Spectral Library. Although GC provides a high chromatographic resolving power, it can only be used for the analysis of volatile components. Therefore, some macromolecules and nonvolatile metabolites cannot be analyzed by GC (Begley et al., 2009; Schauer et al., 2005).

In contrast to gas chromatography, liquid chromatography (LC) does not provide high chromatographic resolution, but can be used to analyze a wider range of compounds. Chromatographic resolving power and peak widths are dependent on the column dimensions (internal diameter and length), stationary phase, mobile phase flow rate, and temperature.

The high resolving power of Ultra Performance Liquid Chromatography (UPLC) is provided by the higher flow rate, increased pressure, and smaller diameter of particles packed in the column (sub-2 μm) (Wang et al., 2011). UPLC-MS can detect thousands of compounds in a given sample using different types of columns. The most commonly used columns are reverse-phase C8 or C18 bonded stationary phases (Wilson et al., 2005). These columns are ideal for relatively non-polar metabolites, including lipids, but they are not applicable for polar metabolites, such as sugars and some amino acids. For these compounds, Hydrophilic Interaction Liquid Chromatography (HILIC) provides better separation of more polar compounds compared to nonpolar lipids, which are practically not retained. Moreover, the combination of both types of separation methods is a feasible approach for sample analysis (Wang et al., 2008).

Capillary Electrophoresis (CE) is the third tool often used for metabolite separation prior to MS detection. CE provides separation efficiency that is equivalent to or better than UPLC

FIG. 3. Mass spectrum of human blood plasma metabolites acquired by DIMS in the positive mode of micrOTOF-Q (BrukerDaltonik Ltd.).

and GC-MS. Similar to all electrophoresis methods, CE is more effective for charged compounds. However, due to technical challenges and limitedness, this technique is not used for metabolite analysis very often (Baidoo et al., 2008; Monton and Soga, 2007).

Data analysis

The presence of a large quantity of metabolites in human blood complicates the reading of biochemical information in metabolic profile. Mass spectra can contain several hundreds or thousands of peaks of metabolites. Therefore, appropriate bioinformatics data processing is essential for the interpretation of metabolic data acquired by MS.

MS data analysis typically requires pre-processing steps, including data alignment and normalization. As several mass spectra are processed, it is important that spectra are aligned and matched for the same metabolite peaks in each sample. Otherwise, it is possible to mistake signals from metabolites with equal molecular masses. Two peaks are considered to be related to the same metabolite if their mass difference does not exceed the defined value (e.g., drift). A drift in the parameters used to identify a specific metabolite is commonly observed in mass spectrometry (retention time, migration time, mass accuracy, and response) and NMR spectroscopy (chemical shift associated with changes in pH or osmolarity) (Barrett, 2012; Beckwith-Hall et al., 1998; Goodacre et al., 2004). There are four commonly used metabolomics normalization methods, including normalization to a constant sum, reference sample (probabilistic quotient normalization), reference feature (creatinine or an internal standard), or sample-specific (dry weight or tissue volume) (Xia et al., 2009).

Selection of the analytical data workflow generally depends on the study objective. For samples classification, a cluster analysis, Principal Component Analysis (PCA), or Independent Component Analysis (ICA) are generally used. In metabolomics, PCA is preferred and can be performed using most statistical software (Dunn et al., 2011). ICA is an

other useful method in metabolomics, as component selection is not as critical and it allows the user to disregard technical variation in MS data, thereby improving the results of the final analysis (Scholz et al., 2004). However, it is often more effective to use several different methods together (Smilde et al., 2005; Vis et al., 2007).

Another frequently used classification method in metabolomics is Partial Least Squares Discriminant Analysis (PLS-DA) (Westerhuis et al., 2008), which is based on the PLS approach (Barker and Rayens, 2003). The standard PLS algorithm can be used, and class labels can be used for the dependent variables. In the two-class case, the dependent variable is usually given a value of 1 for one class and 0 or -1 for the other class. In case of more than 2 classes, dummy variables are defined and a PLS2 algorithm is used. An often used variant of PLS-DA is Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA), where the first components orthogonal to the dependent variable are removed from the data and can be analyzed separately (Trygg and Wold, 2002). This provides a model with a single classification component, while the remaining components describe the other variation that is orthogonal to the class information. OPLS enhances the interpretation of PLS by forcing all classification information into a single component. The prediction power of both models is usually the same, and typical applications of two-class classifications using PLS-DA show a score plot of the classified samples (Trygg and Wold, 2002).

In studies identifying diagnostic biomarkers, the methods involved a learning set of samples where the knowledge about group membership (e.g., diseased or healthy) helps to identify discriminatory biomarkers (Jonsson et al., 2005). One particularly useful visualization method for the qualitative illustration of biomarker utility is the Receiver-Operator Characteristic (ROC) curve (Linden, 2006). The ROC curve is only used in a two-state experimental design (e.g., case-control) and is constructed by plotting the sensitivity versus 1-specificity of a hypothetical decision boundary, moving across the total range of the predictive score (Fig. 4). This plot

FIG. 4. An example of Receiver-Operator Characteristic (ROC) curves for diagnostic tests with high (AUC=0.9) and moderate (AUC=0.65) accuracy. The expert scale of AUC value for the test accuracy rating (Metz, 1978).

includes the points (0.0) and (1.1). If the Area under the ROC Curve (the AUC) is 0.5 (the lower limit), then the variables are distributed equally between the cases and controls and the diagnostic test is useless for discrimination. However, if the area under the ROC curve is 1, then a complete separation of two groups take place, and the samples can be classified with 100% sensitivity (no false negatives) and 100% specificity (no false positives).

The assessment of relative risk of disease development in the presence of some factors is given by an Odds Ratio (OR) (Bland and Altman, 2000; Westergren et al., 2001). The OR indicates the ratio of probabilities of disease development in exposed and unexposed individuals. In the case of metabolic profiling, the peak intensity of a given metabolite or groups of metabolites in the mass spectrum is considered as a factor. Here, the presence or absence of a factor is defined by the peak intensity threshold estimated in accordance with a high value of test accuracy. This can be accomplished using an ROC curve plot by choosing a point with maximal value of sensitivity and specificity. To indicate the reliability of assessment, the Confidence Interval (CI) is calculated, which shows how frequently the observed interval contains the parameter (Morris and Gardner, 1988).

As a result of the recent increase in the number of metabolomics studies, specific software for MS-based metabolomics have been developed (e.g., MET-IDEA (Broeckling et al., 2006), MathDAMP (Baran et al., 2006), and TagFinder (Luedemann et al., 2008)). Many instrument manufacturers offer their own software packages for metabolomics data analysis. Since 2007, Bruker Daltonics Ltd. has provided Metabolic Profiler with their instruments. This commercial software allows for pre-processing of the metabolomics data and comparative analysis of metabolite profiles using raw instrumental data.

Based on the identification of targets or models that are predictive of disease or drug toxicity/efficacy, a greater understanding of the pathophysiological processes involved may be essential. In several studies, the combination of metabolites defined as “biologically interesting” have been studied. These metabolites are usually classified, for example, by the metabolite class or metabolic pathway in databases such as KEGG and HMDB (Denkert et al., 2008; Dunn et al., 2011).

Human Blood Metabolome

Blood is a primary carrier of small molecules in the body and plays an important role in transporting dissolved gases, nutrients, hormones, and metabolic waste. Because blood bathes every tissue and every organ of the body, it essentially serves as a “liquid highway” for all molecules being secreted, excreted, or discarded by the various tissues in response to different physiological needs or stresses. Importantly, tissue lesions, organ dysfunctions, and pathological states can alter both the chemical and protein composition of blood plasma/serum. These facts together with convenient accessibility make blood the most convenient component for clinical testing. Indeed, blood has been the subject of detailed chemical analyses for more than 70 years (Psychogios et al., 2011).

Currently, 10–20 blood metabolites are regularly assessed in the clinical setting. Glucose, blood urea, bilirubin, triglycerides, creatinine, cholesterol, and others are used to diagnose diabetes, renal failure, and other types of diseases. Extensive

tables of normal reference ranges have been published for many blood gases, ions, and at least 100 metabolites (Burtis et al., 2008). In the 2000s, scientific groups began to “globally” characterize the human metabolome. The Human Metabolome Project (HMP) was launched in 2004 as part of an effort to identify and quantify all detectable metabolites in the human body. In addition to the experimental identification and quantification of hundreds of metabolites in different body fluids, this long-term project is also formally tasked with backfilling and validating the information on all previously identified metabolites. The data obtained from this project are freely available as an electronic database called the Human Metabolome Database (HMDB) (Wishart et al., 2007; Wishart et al., 2009). Currently, the HMDB contains more than 7900 human metabolite entries with high ($>1 \mu\text{M}$) and low ($<1 \text{ nM}$) abundance (<http://www.hmdb.ca>). In addition, Psychogios and colleagues (2011) subsequently created the Serum Metabolome Database (SMDB) (<http://www.serummabolome.ca>). This database is a comprehensive, web-accessible resource containing 4229 confirmed and probable serum/plasma compounds with corresponding concentrations and links to their known disease associations, which were revealed or identified by combined experimental and literature mining efforts.

The developers of the SMDB have analyzed 149 serum samples to date from healthy individuals and patients with cardiovascular diseases, using the following five metabolic profiling methods: 1) high-resolution NMR spectroscopy, 2) GC-MS, 3) lipidomics profiling through TLC/GC-MS, 4) lipid mediators by UPLC-MS/MS, and 5) direct injection (DI) MS/MS. In total, 3564 distinct metabolites, including several exogenous compounds, were identified (Fig. 5). Each method was able to detect a different number of serum metabolites. The greatest number of compounds (3381) was identified by lipidomics profiling with TLC/GC-MS, and the lowest number (49) was identified by NMR spectroscopy. The use of more sophisticated or targeted detection and separation protocols (in this case, TLC/GC-MS was generally used for lipid and fatty acid identification) would have likely led to the detection of more compounds. In addition, NMR and GC-MS identified 29 metabolites in common, while NMR, GC-MS, and DIMS identified 15 metabolites in common (14 amino acids and hexose/glucose). Similarly, the same 53 serum metabolites were identified by DIMS and lipidomics profiling (TLC/GC-FID-MS). However, no one compound was identified by all five methods. These differences in metabolite coverage was due to several reasons, including the separation method, instrument sensitivity, targeted vs. nontargeted analysis and stability, solubility, and volatility of compounds.

Psychogios et al. (2011) showed that the LC-MS or DIMS methods may be the best choice for serum metabolomics studies because they both have sufficient sensitivity. DIMS provides a high-throughput system for metabolic profiling because a large number (hundreds or thousands) of samples can be analyzed in a given period of time. The use of DIMS for the large sample sets probably reflects the difficulty in maintaining reproducibility of the LC-MS technique over many weeks or months, where chromatographic drift or small changes in instrument performance can impair the data analysis (Barrett, 2012; Lin et al., 2010).

Because the global metabolome analysis has to provide valid data about numerous metabolites in a sample, the choice

FIG. 5. The overlap of serum metabolites detected by high-resolution NMR spectroscopy and GC-MS, as well as lipidomics profiling by TLC/GC-MS and lipid mediators by UPLC-MS/MS and DIMS in 149 samples from healthy individuals and patients with cardiovascular diseases. Adapted from Psychogios et al., 2011.

of sample preparation strategy is crucial. Blood sample preparation before metabolite profiling is one of the most challenging parts of the analytical procedure. Currently, no universal technique suitable for blood metabolite profiling exists. The need for an unselective sample preparation procedure is dictated by the need to analyze as wide a range of metabolites as possible. This is generally accomplished by using the simplest sample preparation procedures, such as dilution, solvent precipitation, or solvent extraction. More selective procedures, such as liquid–liquid extraction, are usually avoided in metabolomics studies (Tulipani et al., 2013; Vuckovic, 2012). Deproteinization is the minimum sample treatment required prior to MS analysis of blood metabolites and is commonly achieved using organic solvent precipitation (Bruce SJ et al., 2009), protein denaturation, or ultrafiltration (Psychogios et al., 2011). Protein precipitation with an organic solvent is the most commonly used method in the global metabolomics of blood (Vuckovic, 2012). In addition to protein removal, the addition of an organic solvent to biofluid also disrupts any bindings between metabolites and proteins present in solution, and thus allows for an analysis of bound and unbound (free) blood metabolites. Different solvents and their combinations (acetonitrile, acetone, methanol, ethanol, dichloromethane, and others) have been used for the extraction of metabolites from blood-derived samples (plasma and serum). Typically, methanol is preferred over acetonitrile for blood samples, because it produces a flocculent precipitate and not the gummy mass obtained with acetonitrile. Methanol also gives a clearer supernatant and may prevent the drug

entrapment that can be observed after acetonitrile precipitation (Gika and Teodoridis, 2011). Metabolomics studies typically compare the relative levels of metabolites in a large number of samples, and thus reproducibility for a wide range of metabolites is absolutely necessary to ensure that small changes in these metabolite levels can be accurately determined. Metabolite losses are avoided by minimizing the overall number of steps in the procedure, and short sample preparation times facilitate high sample throughput.

Application of Metabolomics in Clinical Studies

Metabolomics plays a major role in greater pathophysiological understanding, but the transfer of metabolomics data to the clinical setting is relatively slow. Recently, metabolomics studies have been used in many fields, such as analyzing the physiologic response to toxicology (Zhang et al., 2012), nutrition (Lee et al., 2010; Wong and Lodge 2012), and cancer (Lokhov et al., 2010; 2012; Sreekumar et al., 2009; Zhou et al., 2012), as well as diabetes (Wang et al., 2011) and disease diagnosis (Sun et al., 2012). The blood metabolome is closely connected with all organs and tissues, and therefore it is convenient and informative for biomarker discovery. Use of metabolomics in the clinical setting is currently focused on discovering new biomarkers for detecting the early stages of a disease.

In recent years, the number of metabolomics studies aimed at discovering new biomarkers in socially significant diseases has increased, including oncology, as well as cardiovascular

and neural diseases (Kaddurah-Daouk and Krishnan, 2009; Lokhov et al., 2010; 2012; Sabatine et al., 2005; Soga et al., 2011; Wang et al., 2011). An NMR metabolic analysis of serum samples from 38 patients with epithelial ovarian cancer, 12 patients with benign ovarian cysts, and 53 healthy women provided a clear distinction of these patients from the healthy controls with regard to their metabolic profile and thus deserves further evaluation as a potential novel strategy for the early detection of epithelial ovarian cancer (Odunsi et al., 2005). Another study assessing the metabolic profile of blood plasma from patients with second stage prostate cancer ($n=40$) and healthy volunteers ($n=30$) revealed six cancer-specific metabolites, which provided a diagnostic accuracy that was higher than the prostate specific antigen (PSA) test (Lokhov et al., 2010). At another study, a comparative metabolomics analysis of blood samples obtained from 36 healthy individuals before and after exercise stress testing found that 18 of them demonstrated inducible ischemia. In addition, six metabolites were found to differentiate the two groups with a high degree of accuracy, including citric acid (Sabatine et al., 2005). Sun et al. (2012) compared the metabolic profiles in serum samples from 17 hospitalized patients with newly diagnosed acute kidney injury with the profiles from age-matched subjects with normal kidney function and observed changes in acylcarnitines, amino acids (methionine, homocysteine, pyroglutamate, asymmetric dimethylarginine, and phenylalanine), arginine, and several lysophosphatidylcholines (Sun et al., 2012). These findings indicate that the novel biomarkers identified by metabolomics analysis can accurately differentiate the patients with these respective diseases from healthy controls. Soga and colleagues (2011) used serum metabolomics to analyze low molecular weight metabolites and discovered noninvasive and reliable biomarkers for the rapid screening and diagnosis of liver diseases. In particular, they found that γ -glutamyl dipeptides are key biomarkers for liver diseases, and different levels of these peptides have the power to distinguish patient with nine types of liver disease from healthy controls.

In addition to the utility as diagnostics, metabolite profiles can predict the development of disease (Lokhov et al., 2012; Wang et al., 2011). Wang and colleagues used metabolomics to predict the development of diabetes. Metabolic profiling was performed on samples from 189 individuals with new-onset diabetes that had developed during a 12-year follow-up period and from 189 propensity-matched controls who did not develop diabetes. Five branched chain and aromatic amino acids were found to have highly significant associations with the future development of diabetes, including isoleucine, leucine, valine, tyrosine, and phenylalanine. In addition, the combination of three amino acids was able to predict future diabetes (Wang et al., 2011). These results underscore the notion that most of these diseases have a multifactorial etiology and the combination of metabolites provides a more powerful prediction of disease than any single factor.

Metabolomics offers the opportunity to investigate global changes in metabolites from biological samples, and indeed there has been substantial interest in this technology from the field of nutritional research (Gibney et al., 2005; Rezzi et al., 2007). There are numerous potential opportunities for these methods in this field, including the assessment of interactions of whole diet or individual nutrients on metabolic pathways, the metabolic interactions between diet and disease, and the development of biomarkers for dietary exposure and/or

disease (Lodge, 2010). Metabolomics technologies have been applied to a number of nutritional studies, allowing investigators to compare the influence of dietary patterns on metabolite profiles, monitor nutrient metabolism, or identify novel metabolic pathways influenced by individual nutrients (Primrose et al., 2011). Wong and Lodge used a nontargeted metabolomics approach to identify changes to the plasma metabolome following vitamin E supplementation in humans (Wong and Lodge, 2012). A PLS-DA model was able to discriminate between samples taken pre- and post-vitamin E supplementation from a relatively homogenous healthy adult male population ($n=10$). Their results suggest that vitamin E supplementation influences phospholipid metabolism and induces lysophosphatidylcholine generation, thereby leading to a general pro-inflammatory response.

Metabolomics studies can provide a greater understanding of disease mechanisms, identify new diagnostic markers, elucidate drug or xenobiotic effects, and increase the ability to predict individual variation in drug response phenotypes (pharmacometabolomics). Using complex technologies, the field of pharmacometabolomics attempts to quantify drug interventions depending on a predose mathematic model of individual metabolic status. The metabolite profile represents all interactions between various aspects of the genome, microbiome (the symbiotic human flora), parasites, environment, and others. Therefore, a metabolite profile is very informative not only for the efficacious dose of a given drug, but also for modifying drug therapy based on patient characteristics, including genetic variations, metabolism, and environmental intervention (Gutiu et al., 2010; Kaddurah-Daouk et al., 2008; Schnackenberg and Beger, 2008).

Comparison of Metabolomics with the Other “Omics”

Until recently, functional genomics was the main “omics” tool for biological objects research. Identifying all gene products simultaneously made it possible to obtain general information on the disease mechanism and to identify suitable therapeutic targets. In some cases this approach provided useful results. However, there are a number of practical and theoretical limits when using mRNA and protein profiling as a universal tool for understanding of disease mechanisms and diagnosis. There is great difficulty in correlating the information encoded at the genome level to a molecular phenotype, because there is often a lack of association between mRNA levels and protein abundance, various protein half-lives, and complex interactions between proteins that give rise to the phenotypic characteristics. Moreover, a transcriptome and proteome analysis is costly and requires multiple repetitions in order to confirm the validity of findings, which reduces its applicability. Therefore, metabolomics is currently the most effective postgenomics technology due to the cost-effective analysis and robustness of information obtained for all processes occurring in an organism. Metabolic data account for the effects of all possible factors (both endogenous and exogenous) influenced in the organism, which allows for a conclusion to be drawn on the mechanism of disease, diagnostic markers, or individual variation in drug response (Kaddurah-Daouk et al., 2008; Schnackenberg and Beger, 2006).

However, despite significant advances, there are several limitations of the current technology that still exist, including the lack of a single method for a comprehensive analysis of the

metabolome, incomplete spectral libraries and databases, and flaws in current software for data extraction and analysis. Thus, further research is still needed before proposing an ideal method for blood metabolite analysis that can replace conventional diagnosis in clinical practice. Future technology development combined with more robust data analysis and bioinformatic tools will help overcome current limitations and fully integrate small molecule biochemistry with systems biology. Because metabolomics is complementary to genomics, transcriptomics, and proteomics, the full integration of these fields is the basis of systems biology, and only their combined application will enable for a full description of a living system at all biomolecular levels (Homuth et al., 2012; Lindon et al., 2006; Richards et al., 2010).

Conclusions

In the present postgenomics era, the significance of metabolic profiling as one of the latest “omics” platforms for biomarker discovery and discerning pathophysiological processes, as well as assessing drug effectiveness and toxicity, has tremendously increased. Thoughtful integration of the findings from metabolic profiling with multidimensional data from genomics, transcriptomics, and proteomics studies can provide comprehensive system level insights regarding the putative disease mechanisms and drug therapy and other health intervention outcomes (e.g., nutrition, vaccines, etc.). Metabolic profiling of blood, which can identify thousands of metabolites at one time, has shown significant results in many scientific and clinic applications to date. Further improvements in the sensitivity and selectivity of analytical techniques and the development and routine use of novel methods are certain to lead to the discovery of novel molecular targets in the near future. However, as with other profiling platforms, including proteomics, transcriptomics, and genomics, which aim to discover clinically relevant biomarkers, the validation of obtained metabolomics results remains important for metabolomics as well. Hence, future metabolomics studies should focus on the standardization, proper experiment design, and validation.

Ultimately, metabolomics studies of peripheral blood offer much promise for postgenomics diagnostics, both by virtue of easy tissue access for biomarker research and the opportunity for integrated multidimensional “omics” biomarkers that are represented at the level of the blood metabolome.

Acknowledgments

This work was supported by the Programme “Proteomics in Medicine and Biotechnology” of the Russian Academy of Medical Sciences and the State Contract no. 16.522.12.2002 with the Ministry of Education and Science of the Russian Federation.

Author Disclosure Statement

The authors declare that no conflicting financial interests exist.

References

- Atherton HJ, Gulston MK, Bailey NJ, et al. (2009). Metabolomics of the interaction between PPAR- α and age in the PPAR- α -null mouse. *Mol Syst Biol* 5, 259.
- Baidoo EE, Benke PI, Neuss C, et al. (2008). Capillary electrophoresis-fourier transform ion cyclotron resonance mass spectrometry for the identification of cationic metabolites via a pH-mediated stacking-transient isotachophoretic method. *Anal Chem* 80, 3112–3122.
- Baran R, Kochi H, Saito N, et al. (2006). MathDAMP: A package for differential analysis of metabolite profiles. *BMC Bioinformatics* 7, 530.
- Barker M, and Rayens W. (2003). Partial least squares for discrimination. *J Chemometr* 17, 166–173.
- Barrett D. (2012). Advances in metabolic profiling. *Bioanalysis* 4, 643–644.
- Beckwith-Hall BM, Nicholson J K, Nicholls AW, et al. (1998). Nuclear magnetic resonance spectroscopic and principal components analysis investigations into biochemical effects of three model hepatotoxins. *Chem Res Toxicol* 11, 260–272.
- Begley P, Francis-McIntyre S, Dunn WB, et al. (2009). Development and performance of a gas chromatography-time-of-flight mass spectrometry analysis for large-scale nontargeted metabolomic studies of human serum. *Anal Chem* 81, 7038–7046.
- Bland JM, and Altman DG. (2000). The odds ratio. *BMJ* 320, 1468.
- Brindle JT, Antti H, Holmes E, et al. (2002). Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using ^1H -NMR-based metabolomics. *Nature Med* 8, 1439–1444.
- Broeckling CD, Reddy IR, Duran AL, Zhao X, and Sumner LW. (2006). MET-IDEA: Data extraction tool for mass spectrometry-based metabolomics. *Anal Chem* 78, 4334–4341.
- Bruce SJ, Tavazzi I, Parisod V, Rezzi S, Kochhar S, and Guy PA. (2009). Investigation of human blood plasma sample preparation for performing metabolomics using ultrahigh performance liquid chromatography/mass spectrometry. *Anal Chem* 81, 3285–3296.
- Burtis CA, Ashwood AE, Bruns DE, and Tietz NW. (2008). *Tietz Fundamentals of Clinical Chemistry*. Saunders Elsevier, St. Louis, Mo.
- Denkert C, Budczies J, Weichert W, et al. (2008). Metabolite profiling of human colon carcinoma—Deregulation of TCA cycle and amino acid turnover. *Mol Cancer* 7, 72.
- Dunn WB, Broadhurst DI, Atherton HJ, Goodacre R, and Griffin JL. (2011). Systems level studies of mammalian metabolomes: The roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem Soc Rev* 40, 387–426.
- Ellis DI, Dunn WB, Griffin JL, Allwood JW, and Goodacre R. (2007). Metabolic fingerprinting as a diagnostic tool. *Pharmacogenomics* 8, 1243–1266.
- Gibney MJ, Walsh M, Brennan L, Roche HM, German B, and van Ommen B. (2005). Metabolomics in human nutrition: Opportunities and challenges. *Am J Clin Nutr* 82, 497–503.
- Gika H, and Teodoridis G. (2011). Sample preparation prior to LC-MS-based metabolomics/metabonomics of blood derived samples. *Bioanalysis* 3, 1647–1661.
- Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, and Kell DB. (2004). Metabolomics by numbers: Acquiring and understanding global metabolite data. *Trends Biotechnol* 22, 245–252.
- Gutiu IA, Andries A, Radulescu F, Georgescu A-M, and Cioaca D. (2010). Pharmacometabolomics, pharmacogenomics and personalized medicina. *Romanian J Int Med* 48, 187–191.
- Homuth G, Teumer a, Volker U, and Nauck M. (2012). A description of large-scale metabolomics studies: Increasing value by combining metabolomics with genome-wide SNP genotyping and transcriptional profiling. *J Endocrinol* 215, 7–28.

- Jonsson P, Bruce SJ, Moritz T, Trygg J, Sjöström M, Plumb R, et al. (2005). Extraction, interpretation and validation of information for comparing samples in metabolic LC/MS data sets. *Analyst* 130, 701–707.
- Kaddurah-Daouk R, and Krishnan KR. (2009). Metabolomics: A global biochemical approach to the study of central nervous system diseases. *Neuropsychopharmacology* 34, 173–186.
- Kaddurah-Daouk R, Kristal BS, and Weinshilboum RM. (2008). Metabolomics: A global biochemical approach to drug response and disease. *Ann Rev Pharmacol Toxicol* 48, 653–683.
- Kell DB, and Oliver SG. (2004). Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *BioEssays* 26, 99–105.
- Kirschenlohr HL, Griffin JL, Clarke SC, et al. (2006). Proton NMR analysis of plasma is a weak predictor of coronary artery disease. *Nature Med* 12, 705–710.
- Lee R, West D, Phillips SM, and Britz-McKibbin P. (2010). Differential metabolomics for quantitative assessment of oxidative stress with strenuous exercise and nutritional intervention: Thiol-specific regulation of cellular metabolism with N-acetyl-L-cysteine pretreatment. *Anal Chem* 82, 2959–2968.
- Lin L, Yu Q, Yan X, et al. (2010). Direct infusion mass spectrometry or liquid chromatography mass spectrometry for human metabolomics? A serum metabolomic study of kidney cancer. *Analyst* 135, 2970–2978.
- Linden A. (2006). Measuring diagnostic and predictive accuracy in disease management: An introduction to receiver operating characteristic (ROC) analysis. *J Eval Clin Pract* 12, 132–139.
- Lindon JC, Holmes E, and Nicholson JK. (2006). Metabolomics techniques and applications to pharmaceutical research & development. *Pharmaceut Res* 23, 1075–1088.
- Lodge JK. (2010). Symposium 2: Modern approaches to nutritional research challenges: Targeted and non-targeted approaches for metabolite profiling in nutritional research. *Proc Nutr Soc* 69, 95–102.
- Lokhov PG, Dashtiev MI, Moshkovskii SA, and Archakov AI. (2010). Metabolite profiling of blood plasma of patients with prostate cancer. *Metabolomics* 6, 156–163.
- Lokhov PG, Kharybin ON, and Archakov AI. (2012). Diagnosis of lung cancer based on direct-infusion electrospray mass spectrometry of blood plasma metabolites. *Intl J Mass Spectrom* 309, 200–205.
- Lokhov PG, Trifonova OP, Maslov DL, and Archakov AI. (2013). Blood plasma metabolites and the risk of development of lung cancer in Russia. *Eur J Cancer Prevent* 22, 335–341.
- Luedemann A, Strassburg K, Erban A, and Kopka J. (2008). TagFinder for the quantitative analysis of gas chromatography–mass spectrometry (GC-MS)-based metabolite profiling experiments. *Bioinformatics* 24, 732–737.
- Metz CE. (1978). Basic principles of ROC analysis. *Sem Nuclear Med* 8, 283–298.
- Monton MR, and Soga T. (2007). Metabolome analysis by capillary electrophoresis-mass spectrometry. *J Chromatog A* 1168, 237–246; discussion 236.
- Morris JA, and Gardner MJ. (1988). Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. *Br Med J (Clin Res Ed)* 296, 1313–1316.
- Nicholson JK, and Lindon JC. (2008). Systems biology: Metabolomics. *Nature* 455, 1054–1056.
- Oduksi K, Wollman RM, Ambrosone CB, et al. (2005). Detection of epithelial ovarian cancer using ¹H-NMR-based metabolomics. *Intl J Cancer* 113, 782–788.
- Primrose S, Draper J, Elsom R, et al. (2011). Metabolomics and human nutrition. *Br J Nutr* 105, 1277–1283.
- Psychogios N, Hau DD, Peng J, et al. (2011). The human serum metabolome. *PLoS One* 6, e16957.
- Rezzi S, Ramadan Z, Fay LB, and Kochhar S. (2007). Nutritional metabolomics: Applications and perspectives. *J Proteome Res* 6, 513–525.
- Richards SE, Dumas ME, Fonville JM, Ebbels TMD, Holmes E, and Nicholson JK. (2010). Intra- and inter-omic fusion of metabolic profiling data in a systems biology framework. *Chemomet Intell Lab Systems* 104, 121–131.
- Rooney OM, Troke J, Nicholson JK, and Griffin JL. (2003). High-resolution diffusion and relaxation-edited magic angle spinning ¹H NMR spectroscopy of intact liver tissue. *Magn Res Med* 50, 925–930.
- Sabatine MS, Liu E, Morrow DA, et al. (2005). Metabolomic identification of novel biomarkers of myocardial ischemia. *Circulation* 112, 3868–3875.
- Schauer N, Steinhauser D, Strelkov S, et al. (2005). GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Lett* 579, 1332–1337.
- Schnackenberg LK. (2007). Global metabolic profiling and its role in systems biology to advance personalized medicine in the 21st century. *Expert Rev Mol Diagn* 7, 247–259.
- Schnackenberg LK, and Beger RD. (2006). Monitoring the health to disease continuum with global metabolic profiling and systems biology. *Pharmacogenomics* 7, 1077–1086.
- Schnackenberg LK, and Beger RD. (2008). The role of metabolic biomarkers in drug toxicity studies. *Toxicol Mechan Methods* 18, 301–311.
- Scholz M, Gatzek S, Sterling A, Fiehn O, and Selbig J. (2004). Metabolite fingerprinting: Detecting biological features by independent component analysis. *Bioinformatics* 20, 2447–2454.
- Smilde AK, Jansen JJ, Hoefsloot HC, Lamers RJ, van der Greef J, and Timmerman ME. (2005). ANOVA-simultaneous component analysis (ASCA): A new tool for analyzing designed metabolomics data. *Bioinformatics* 21, 3043–3048.
- Smith LM, Maher AD, Cloarec O, et al. (2007). Statistical correlation and projection methods for improved information recovery from diffusion-edited NMR spectra of biological samples. *Anal Chem* 79, 5682–5689.
- Soga T, Sugimoto M, Honma M, et al. (2011). Serum metabolomics reveals γ -glutamyl dipeptides as biomarkers for discrimination among different forms of liver disease. *J Hepatol* 55, 896–905.
- Sreekumar A, Poisson LM, Rajendiran TM, et al. (2009). Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* 457, 910–914.
- Stringer KA, Serkova NJ, Karnovsky A, Guire K, Paine R, 3rd, and Standiford TJ. (2011). Metabolic consequences of sepsis-induced acute lung injury revealed by plasma (1)H-nuclear magnetic resonance quantitative metabolomics and computational analysis. *Am J Physiol Lung Cell Mol Physiol* 300, L4–L11.
- Sun J, Ni B, Zhang A, Wang M, Dong H and Wang X (2012). Serum metabolomic profiles from patients with acute kidney injury: A pilot study. *Journal of Chromatography B Analyt Technol Biomed Life Sci* 893–894, 107–113.
- Trygg J, and Wold S. (2002). Orthogonal projections to latent structures (O-PLS). *J Chemometrics* 16, 119–128.
- Tulipani S, Llorach R, Urpi-Sarda M, and Andres-Lacueva C. (2013). Comparative analysis of sample preparation methods to handle the complexity of the blood fluid metabolome: when less is more. *Anal Chem* 85, 341–348.

- Vaidyanathan S, Kell DB, and Goodacre R. (2002). Flow-injection electrospray ionization mass spectrometry of crude cell extracts for high-throughput bacterial identification. *J Am Soc Mass Spectrom* 13, 118–128.
- Vis DJ, Westerhuis JA, Smilde AK, and van der Greef J. (2007). Statistical validation of megavariable effects in ASCA. *BMC Bioinformatics* 8, 322.
- Vuckovic D. (2012). Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. *Anal Bioanal Chem* 403, 1543–1548.
- Wang TJ, Larson MG, Vasan RS, et al. (2011). Metabolite profiles and the risk of developing diabetes. *Nature Med* 17, 448–453.
- Wang X, Sun H, Zhang A, Wang P, and Han Y. (2011). Ultra-performance liquid chromatography coupled to mass spectrometry as a sensitive and powerful technology for metabolomic studies. *J Separ Sci* 34, 3451–3459.
- Wang Y, Lehmann R, Lu X, Zhao X, and Xu G. (2008). Novel, fully automatic hydrophilic interaction/reversed-phase column-switching high-performance liquid chromatographic system for the complementary analysis of polar and apolar compounds in complex samples. *J Chromatog A* 1204, 28–34.
- Westergren A, Karlsson S, Andersson P, Ohlsson O, and Hallberg IR. (2001). Eating difficulties, need for assisted eating, nutritional status and pressure ulcers in patients admitted for stroke rehabilitation. *J Clin Nurs* 10, 257–269.
- Westerhuis JA, Hoefsloot HCJ, Smit S, et al. (2008). Assessment of PLSDA cross validation. *Metabolomics* 4, 81–89.
- Wilson ID, Nicholson JK, Castro-Perez J, et al. (2005). High resolution “ultra performance” liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies. *J Proteome Res* 4, 591–598.
- Wishart DS, Tzur D, Knox C, et al. (2007). HMDB: the Human Metabolome Database. *Nucleic Acids Res* 35 (Database issue), D521–526.
- Wishart DS, Knox C, Guo AC, et al. (2009). HMDB: A knowledgebase for the human metabolome. *Nucleic Acids Res* 37 (Database issue), D603–610.
- Wong M, and Lodge JK. (2012). A metabolomic investigation of the effects of vitamin E supplementation in humans. *Nutrit Metabol* 9, 110.
- Xia J, Psychogios N, Young N, and Wishart DS. (2009). Meta-boAnalyst: A web server for metabolomic data analysis and interpretation. *Nucleic Acids Res* 37, W652–660.
- Zhang A, Sun H, Wang P, Han Y, and Wang X. (2012). Metabonomics for discovering biomarkers of hepatotoxicity and nephrotoxicity. *Pharmazie* 67, 99–105.
- Zhang S, Nagana Gowda GA, Ye T, and Raftery D. (2010). Advances in NMR-based biofluid analysis and metabolite profiling. *Analyst* 135, 1490–1498.
- Zhou L, Wang Q, Yin P, et al. (2012). Serum metabolomics reveals the deregulation of fatty acids metabolism in hepatocellular carcinoma and chronic liver diseases. *Anal Bioanal Chem* 403, 203–213.

Address correspondence to:

Dr. Oxana Trifonova
Institute of Biomedical Chemistry RAMS
Pogodinskaya Street, 10
119121 Moscow
Russia

E-mail: oxana.trifonova@gmail.com